ENVIRONMENTAL HEALTH PERSPECTIVES
http://www.ehponline.org

Fine-Scale Exposure to Allergenic Pollen in the Urban Environment: Evaluation of Land Use Regression Approach

Jan Hjort, Timo T. Hugg, Harri Antikainen, Jarmo Rusanen, Mikhail Sofiev, Jaakko Kukkonen, Maritta S. Jaakkola, and Jouni J.K. Jaakkola

http://dx.doi.org/10.1289/ehp.1509761

Received: 27 January 2015
Accepted: 5 October 2015
Advance Publication: 9 October 2015

Environ Health Perspect DOI: 10.1289/ehp.1509761
Advance Publication: Not Copyedited

Fine-Scale Exposure to Allergenic Pollen in the Urban Environment: Evaluation of Land Use Regression Approach

Jan Hjort,[1] Timo T. Hugg,[2] Harri Antikainen,[1] Jarmo Rusanen,[1] Mikhail Sofiev,[3] Jaakko Kukkonen,[3] Maritta S. Jaakkola,[2] and Jouni J.K. Jaakkola[2]

[1]Department of Geography, University of Oulu, Oulu, Finland; [2]Center for Environmental and Respiratory Health Research, University of Oulu, Oulu, Finland; [3]Finnish Meteorological Institute, Helsinki, Finland

Address correspondence to J. Hjort, Department of Geography, University of Oulu, P.O. Box 3000, FI-90014 University of Oulu, Oulu, Finland. Telephone: 358 (0)29 4481704. E-mail: jan.hjort@oulu.fi

Short running title: Evaluation of LUR for pollen exposure assessment

Acknowledgments: This research was funded by the Research Council for Health, the Academy of Finland grants no. 266314, 267675 and 267995 (APTA Consortium) and the University of Oulu Strategic Funding. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The constructive comments of three anonymous reviewers are acknowledged.

Competing financial interest: The authors declare they have no actual or potential competing financial interests.

Abstract

Background: Despite the recent developments in physically- and chemically- based analysis of atmospheric particles, no models exist for resolving the spatial variability of pollen concentration at urban scale.

Objectives: We developed land use regression (LUR) approach for predicting spatial fine-scale allergenic pollen concentrations in the Helsinki Metropolitan Area, Finland, and evaluated the performance of the models against available empirical data.

Methods: We utilized grass pollen data monitored at 16 sites in an urban area during the peak pollen season and geospatial environmental data. The main statistical method was generalized linear model (GLM).

Results: GLM based LURs explained 79% of the spatial variation in the grass pollen data based on all samples, and 47% of the variation when samples from two sites with very high concentrations were excluded. In model evaluation, prediction errors ranged from 6% to 26% of the observed range of grass pollen concentrations. Our findings support the use of geospatial data-based statistical models to predict the spatial variation of allergenic grass pollen concentrations at intra-urban scales. A remote sensing based vegetation index was the strongest predictor of pollen concentrations for exposure assessments at local scales.

Conclusions: LUR approach provides new opportunities to estimate the relations between environmental determinants and allergenic pollen concentration in human-modified environments at fine spatial scales. The presented approach could potentially be applied to estimate retrospectively pollen concentrations to be used for long-term exposure assessments.

Introduction

Asthma is globally the most common chronic disease in children and it affects approximately 7.7% of the working-age population in the United States (Henneberger et al. 2011) and 3–9% of adults in Finland and in other parts of Europe (Boutin-Forzano et al. 2007; Jaakkola et al. 2002; Pallasaho et al. 2011). Allergies are even more common with the prevalence of 20–40% among children (Aït-Khaled et al. 2009; Asher et al. 2006). Emerging climate change will influence temperature, precipitation, and the spatial distribution of pollen species with strong allergenic properties, and these changes may have profound effects on both the etiology of asthma and allergies and the occurrence of symptoms among subjects with these diseases (e.g. Beggs and Bambrick 2005; Cecchi et al. 2010; D'Amato et al. 2014; Gilmour et al. 2006).

Grass (*Poaceae*) pollen is the most widespread group of pollen allergens worldwide, and is the most frequent cause of pollen allergy in Europe and one of the most common causes in the U.S. (D'amato et al. 2007; White and Bernstein 2003). In general, the majority of individuals suffering from allergic rhinitis experience seasonal pollen-related symptoms (Blomme et al. 2013). Pollen allergy has been identified in 80–90% of children with asthma and 40–50% of adults with asthma (Taylor et al. 2007). The importance of pollen is highlighted in urban environments where the prevalence of allergy has been estimated to be higher compared to rural environments (e.g. Majkowska-Wojciechowska et al. 2007; Priftis et al. 2007).

Conditions that promote plant growth and reproduction, such as higher temperatures and CO_2 concentrations (Singer et al. 2005; Ziska et al. 2004) have showed to increase pollen production and earlier start of pollen season in urban areas compared with surrounding rural areas (Ziska et al. 2003). Importantly, there is evidence that allergenic potential of polluted (urban) pollen is stronger than that of non-polluted pollen grains (Aina et al. 2010; Majd et al. 2004). Thus, relatively more favorable vegetation growth conditions, higher pollen counts, earlier start of pollen season and greater allergenic potency may have a pronounced effect on the occurrence and the severity of allergic symptoms, and the risk of asthma and allergies in urban environments (Ziska et al. 2003; Beggs 2004; Gilmour et al. 2006). The global trend of urbanization and formation of mega-cities will increase the human exposure to mixtures of chemical and biological pollutants (WHO 2013).

Previous epidemiologic studies have predominantly assessed exposure on the basis of pollen data from only one or few monitoring stations (Caillaud et al. 2014; Delfino et al. 2002). An additional limitation in several studies is that pollen data have been based on roof-level measurements (e.g., at the heights of 10–30 m) and thus do not reflect properly the most common exposure at breathing level (e.g. 1.5 m). Roof-level samplers collect particles from larger geographic area, reflecting mainly the influences of regional sources of pollen grains (O'Rourke and Lebowitz 1984). Thus, roof-level data do not address intra-urban spatial heterogeneity of pollen concentrations and they provide less accurate data for exposure assessment when studying potential human health effects (Peel at al. 2013). Consequently, there is an urgent need for studies which can quantify the relations between environmental determinants and allergenic pollen concentrations across urban gradient at local (≤100–300 m) scales (Haberle et al. 2014; WHO 2013).

In recent years, there has been a significant progress in the modeling of transportation and concentration of atmospheric pollen at large spatial scales (Kukkonen et al. 2012; Skjøth et al. 2009; Sofiev and Bergman 2012; Sofiev et al. 2012). However, even the most sophisticated current methods cannot estimate the concentrations of pollens at fine spatial scale. In this context, a land use regression (LUR) approach based on readily available geographic information system (GIS) data could provide a cost-efficient and reasonably accurate method for predicting the variability of pollen concentrations at fine spatial scales (Beelen et al. 2013; Richardson et al. 2013).

The overall objective of this study was to improve and evaluate the LUR methodology for the prediction of spatial variability of allergenic pollen concentrations in the urban area. We focused on grasses, as one of the most important allergenic plant groups. More precisely, we addressed the following specific objectives: (i) to model intra-urban spatial variation of grass pollen concentrations and (ii) to determine the best environmental predictors of intra-urban variation in grass pollen concentrations.

Methods

Study area

The study was conducted in the Helsinki Metropolitan Area (1.1 million inhabitants), Southern Finland (60°10′15″N, 24°56′15″E) (Fig. 1). The study area includes constructed urban environments with limited amount of vegetation and natural environments covered by diverse vegetation. The study area has the characteristics of both maritime and continental climates. The mean annual temperature is +5.9˚C and the mean annual precipitation is 655 mm (1981–2010; Pirinen et al. 2012). The area belongs to the temperate coniferous-mixed forest zone. Grasses are typical pioneer plant groups (i.e. those which are the first to colonize newly exposed land surfaces) in the ground layer of the area.

Data collection

Grass pollen data

Two sampling lines, each with a length of three kilometers, were placed within the cities of Helsinki and Espoo (Fig. 1). The selected sampling sites differed from each other with respect to land use and vegetation type. These two categories of urban environments (urban environment in central Helsinki, and residential suburban area in Espoo) were selected for evaluation because grass pollen exposures in these areas may be representative of the exposures experienced by many urban populations. Altogether, pollen grains were monitored at 16 different sites (1.5 m above ground surface) during the grass pollen season in 2013. The sampling was conducted daily (except on rainy days) from 27 June through 21 July 2013, during the peak grass pollen season. See Supplemental Material, Table S1 for a summary of weather conditions during the sampling period.

Rotorod-type samplers were used for pollen monitoring (Rantio-Lehtimäki et al. 1992). To minimize problems with oversampling (Sterling and Lewis 1998) samples were collected at each site for only 30 minutes each morning (between 8:00 and 11:30) and each afternoon (between 13:00 and 16:30). On each day the specific sampling time differed among the individual sites, and the specific sampling times at each site differed from day to day. In addition, sampling days alternated between the eight Helsinki sites and the eight Espoo sites (Figure 1). For additional details, see Supplemental Material, Sampling of grass pollen data. Pollen measurements were converted into volumetric equivalents expressed as the concentration of pollen grains per cubic meter of air sampled (grains m^{-3}). The grass pollen data were subdivided into three subsets that each represented the average concentration of pollens at the breathing zone (1.5 m) at each one of the 16 sampling sites during a two week period. The first was used to calibrate the LUR model, and included data collected June 27–July 9 during the afternoon sampling period (13:00–16:30). The second and third were used for independent evaluation of the model, with the second collected June 27–July 9 during the morning sampling period (8:00–11:30), and the third collected July 10–21 during the morning and afternoon sampling period. In the calibration dataset, average pollen concentrations for 2 of the 16 sites were very high (120.2 and 1758.0 grains m^{-3}) compared with average concentrations for the other 14 sites (1.4–23.9 grains m^{-3}). Therefore, in addition to analyzing the data from all 16 sites (hereafter referred to as the n = 16 dataset), we performed a second set of analyses after excluding the data from the two sites with very high concentrations (hereafter referred to as the n = 14 dataset) in order to assess the robustness of our model with regard to potential outliers.

Environmental determinants

Altogether eight GIS-based geospatial environmental determinants (Table 1) were computed at 25, 50, 100, 300, 500 and 1000 m buffer sizes (Lovasi et al. 2013). The range of buffer sizes was selected based on the knowledge of dispersal of grass pollens (e.g. Skjøth et al. 2013) and the properties of geospatial datasets (e.g. accuracy and resolution). The applied determinants included two remote sensing-based indices [tasseled cap transformation (TC) Brightness and TC Greenness (Lillesand et al. 2004)], four land use determinants (wasteland, park, field, and urban land use) compiled from the 2010 SLICES land use classification (SLICES 2011) and two land use/cover variables (deciduous forest and mixed forest) computed using the 2006 CORINE land cover database (Finnish Environmental Institute 2009). The values of the environmental determinants were computed using ArcGIS 10.2 with buffers at various sizes.

The source of the remote sensing data was selected based on the following criteria: the images should be freely available, they should have global coverage with high spatial resolution (10–50 m), and the data should be available for several decades (to facilitate linking this data to long-term epidemiologic cohort data; e.g. Lovasi et al. 2013) (Fig. 2). Consequently, the remote sensing-based indices, i.e. TC brightness and TC greenness (Lillesand et al. 2004), were computed from orthorectified Landsat TM5 satellite image (considering the acquisition day and cloud cover the most suitable image was from July 14, 2010) using Erdas IMAGINE. We selected the TC greenness variable instead of the more commonly used normalized difference vegetation index (NDVI) (e.g. Skjøth et al. 2013) based on a preliminary analysis which indicated that the TC greenness variable was more highly correlated with grass pollen concentrations [the highest Spearman's rank order correlation coefficients (r_s) = 0.792, $p < 0.001$)] than the NDVI (the highest $r_s = 0.788$, $p < 0.001$).

Statistical analysis

Before the multivariate statistical analyses the distributions of the grass variables were normalized using logarithmic ($\log_{10}$) transformation (normality was statistically confirmed applying Kolmogorov-Smirnov test) (Sokal and Rohlf 1995). The statistical analyses were conducted in three steps to explore the explanation and prediction ability of the selected environmental determinants at various scales. The main statistical methods applied were Spearman's rank correlation analysis (Sokal and Rohlf 1995), hierarchical partitioning (HP; Chevan and Sutherland 1991) and generalized linear modeling (GLM; McCullagh and Nelder 1989). First, r_s between grass pollen concentrations and environmental determinants were calculated to select the optimal buffer sizes. The selection of the optimal buffer size was based on the highest correlation coefficients, requiring also expected sign (e.g. a positive sign for variables describing potential environments for grasses). The correlations were calculated with IBM SPSS Statistics 19 software. Second, to explore the potential effects of multicollinearity in statistical analyses (i.e. intercorrelation among environmental determinants) we applied HP (Chevan and Sutherland 1991) as described in detail in Supplemental Material, Hierarchical partitioning. Third, GLM was employed to model the observed grass pollen concentration differences in the study area. The calibration of the GLM was performed utilizing the standard *glm* function in R. The GLM optimization was based on the forward-selection approach and Akaike's Information Criterion (AIC) (Burnham and Anderson 2004). We evaluated the calibrated model on the basis of the model fit to the data, normality of the residuals (applying the Kolmogorov-Smirnov test), homoscedasticity, spatial independency (applying the Moran's I) and leverage (Beelen et al. 2013; Sokal and Rohlf 1995).

The predictive ability of the final GLMs was assessed using leave-one-out cross validation (CV) of models based on the calibration dataset. In addition, we performed an external validation by comparing calibration model predictions with the two evaluation datasets based on samples collected June 27–July 9 during the morning, and samples collected July 10–21 during the morning and afternoon, respectively. In addition, we visually compared spatial variation in the predicted concentrations with land use and land cover characteristics across the Helsinki Metropolitan Area.

Results

Optimization of the buffer sizes

Environmental variables with the highest Spearman's rank order correlations with grass pollen concentrations based on the full ($n = 16$) dataset were TC greenness (optimum buffer size = 50 m, $r_s = 0.79$, $p < 0.001$), wasteland (300 m, $r_s = 0.73$, $p < 0.001$), urban land use (300 m, $r_s = -0.72$, $p < 0.01$), field (1000 m, $r_s = 0.67$, $p < 0.01$), and deciduous forest (300 m, $r_s = 0.63$, $p < 0.01$) (Supplemental Material, Table S2). For the reduced ($n = 14$) dataset, grass pollen concentrations correlated highest with TC greenness (50 m, $r_s = 0.69$, $p < 0.01$), urban land use (1000 m, $r_s = -0.62$, $p < 0.05$), wasteland (500, $r_s = 0.62$, $p < 0.05$), and mixed forest (50 m, $r_s = 0.62$, $p < 0.05$) (Supplemental Material, Table S3). Thus, the correlation patterns were rather similar for the two datasets, although correlations were weaker for the n = 14 dataset than for the full dataset.

Hierarchical partitioning

The results of the HP analysis i.e. the independent contribution of the environmental determinants at the optimum buffer size are presented in Supplemental Material, Figure S1. The contributions of the TC greenness and urban land use variables were over 15% in both datasets (ranged from 17% to 30%). Moreover, the wasteland variable contributed 22% in the n = 16 dataset and the mixed forest variable 21% in the n = 14 dataset. Due to the small sample sizes only the TC greenness variable was statistically significant ($p < 0.001$) in the larger data set, and none of the determinants were significant ($p > 0.05$) in the smaller dataset.

Calibration and evaluation of generalized linear models

The final GLMs for the n = 16 and n = 14 datasets both included the TC greenness variable, while the n = 16 model also included the wasteland variable. The two GLMs explained 79% and 47% of the variation in the grass pollen data, respectively (D^2 values, Table 2). Based on the exploration of residuals, the assumptions of normal errors and independency were not statistically violated. The larger dataset included one potential outlier (Cook's distance > 1), but when an alternative GLM was calibrated without the outlier observation, the normality and homoscedasticity of the residuals was reduced, and a new potential outlier was identified (data not shown).

Based on the CV results, the prediction errors were 412.9 grains m^{-3} and 5.9 grains m^{-3} for the n = 16 and n = 14 datasets, respectively. However, it should be noted that the mean prediction errors were 6.4% (n = 16) and 19.5% (n = 14) and root-mean-square errors were 23.5% (n = 16) and 26.4% (n = 14) of the observed range of grass pollen concentrations. The calibrated model for the full dataset predicted the measured concentrations of samples collected during the morning (instead of the afternoon) on the same days better than it predicted measured concentrations during the second observation period (July 10–21) (Pearson's $r = 0.84$, $p < 0.001$ and 0.55, $p < 0.05$, respectively) (Figure 3A and B). In contrast, predictions based on the n = 14 dataset were slightly better for the second observation period than for the morning samples ($r = 0.58$, $p < 0.05$ and 0.38, $p > 0.05$, respectively) (Figure 3C and D).

The overall view of the predicted concentrations of grass pollen follows well the land use patterns in the Helsinki Metropolitan Area (Fig. 4). For example, the major source areas of pollens (e.g. grasslands and open semi-natural land use types, Fig. 4A and D) and areas of intensive land use (e.g. built-up and traffic areas, Fig. 4B–D) can be identified. Moreover, that the most densely vegetated environments (i.e. forest, Fig. 4A) are not the areas with the highest concentrations of grass pollens. This result is to be expected, because grasses do not flourish in forests with closed canopy. On the contrary, the models appear to predict too high grass pollen concentrations for some of the cultivated fields (Fig. 4A and B) and managed grass areas (e.g. golf courses). However, the models predict realistically rather low concentrations for extensively managed grass areas, for example around the airport runways in the middle part of the city of Vantaa (Fig. 4).

Discussion

Atmospheric concentration of allergenic pollen is an important public health indicator, the potential of which has not been fully utilized (Ring et al. 2014; WHO 2013). Although the role of pollen in the onset of allergic disease is not well established, exposure to pollen generates symptoms among asthmatics and allergic individuals (Gilmour et al. 2006; Zeldin et al. 2006). Thus, improved prediction of the variability of allergenic pollen concentrations at local scales would be valuable when studying the role of pollen exposure in the development of allergic diseases and sensitization, as well as exacerbations of symptoms among subjects with asthma and allergies. This would also pave the way for improved predictions of the future health impact of climate change through changes in the generation and distribution of pollen.

The LUR models developed in this study are designed for predicting the variability of the breathing-zone level concentrations of allergenic grass pollens in urban environment with very high resolution using multi-source geospatial data. First, this is of importance because physically-based dispersion models are not, at least yet, applicable in the exploration of fine-scale spatial variation of pollen concentrations (Bergman and Sofiev 2012; Sofiev et al. 2012). The main application of the suggested methodology is to assess longer term (from weekly to seasonal) exposures to pollens and not short-term, high-dose exposures (Fig. 2). Second, the measured grass pollen concentrations presented significant intra-urban variation. Consequently, it is necessary to consider also local-scale sources in addition to regional-scale sources or long-distance atmospheric transport when assessing population exposure to grass pollens in urban environments.

Implications for the assessment of exposure and health effects

Associations between pollen concentrations measured by regional pollen monitoring stations and allergic and asthmatic symptoms and lung function variation have been reported (Caillaud et al. 2014; Delfino et al. 2002). However, the role of long-term exposure to allergenic pollen in the development of asthma and other allergic diseases has not been studied, in part because of the lack of appropriate methods for exposure assessment.

Our findings suggest that allergenic grass pollen concentrations can be estimated with reasonable accuracy using geospatial data variables. Because similar data have been collected for several decades, it may be possible to adapt our method to perform retrospective life-time exposure assessment (cf. Gulliver et al. 2013) for cohort members based on historical land use data, pollen data, and individual residential history data (Richardson et al. 2013) (Fig. 2). In this case the local variations in pollen concentration are predicted using past environmental conditions (e.g. Landsat satellite images that are available for over three decades) and the regional level of concentration is determined by a permanent, long-term pollen collector located in the study area (e.g. daily pollen observations from the Helsinki Metropolitan Area are available since the mid-1970). The presented modeling approach could also be applied in cross-sectional studies comparing the prevalence of asthma-related and allergic symptoms according to residential area or time-activity pattern.

One notable issue was the moderate prediction bias when the models were extrapolated to other periods (i.e. under-prediction of grass pollen concentrations when pollen production of the extrapolation period was lower than in the calibration period). However, this potential bias can be taken into account in retrospective predictions, if the overall concentration differences between different periods are known. The long-term pollen concentrations measurements from regional monitoring stations can be used for this purpose.

Data and methodological issues

Our findings reinforce the need to identify optimal buffer sizes for each environmental determinant, as appropriate buffers may vary substantially. In addition, our results support the hypotheses that there are large intra-urban variations in grass pollen concentrations and these heterogeneities are associated to local-scale variations in land use (Skjøth et al. 2013). The remote sensing-based TC greenness variable outperformed land use and land cover variables in this study. In general, remote sensing-based indices have several advantages over the traditional land use variables. Remote sensing data can be acquired over extensive areas and from numerous sources at various spatial, temporal, spectral and radiometric resolutions (Lillesand et al. 2004). More importantly, remote sensing-based variables can be computed to describe the environmental conditions continuously (i.e. each pixel has a continuous numerical value).

In contrast, the land use variables are dichotomized (i.e. the class is present or not) (Katz and Carey 2014). This property impedes the use of distinct land use/cover classes in predictive modeling settings, because the model will predict equal concentrations in areas without variation in the explanatory variable(s). For example, the wasteland variable describing unmanaged grasslands was, in theory, an ideal variable reflecting the sources of grass pollens. However, the amount of this land use type was low in the study area (Fig. 1).

Some data-related and statistical issues should be considered. Firstly, the number of observation sites (n = 16) was limited, which caused challenges in model calibration and evaluation. Robust statistical models usually require dozens of observations (for example, Hjort et al. 2011; Beelen et al. 2013), and hundreds of sample sites may be optimal in analyzing ecological response variables (Stockwell and Peterson 2002). The use of passive aerosol samplers would make it feasible to acquire data from more sites and thus improve the assessment and prediction of spatial variation in pollen concentrations in intra-urban areas (e.g. Hofmann et al. 2014).

Secondly, variables describing meteorological conditions were not included into the set of environmental determinants. The primary goal of our model was to predict the spatial variability of pollen concentrations within an urban area: additional data and a longer sampling period would be required to accurately predict the absolute values of the pollen concentrations. The weather conditions were partially taken into account in the sampling of pollen data (no sampling during rainfall, only daytime sampling) and during the preparation of calibration and evaluation datasets (Supplemental Material, Table S1). Implicitly, the regular effects of meteorological factors in different land use classes are accounted for during the calibration step. We assumed that all other effects of meteorology would be consistent across the study area, which would make them irrelevant for predicting spatial variation. Moreover, we used data averaged over two-week periods (June 27–July 9 and July 10–21, respectively) to reduce the impact of short-term temporal variability.

Thirdly, in addition to the sample size related problems, regression-based statistical methods include several data-based assumptions (Sokal and Rohlf 1995). For example, assumptions of normality, homoscedasticity and independence of errors are often violated in analyzing complex responses. In this study, we aimed at minimizing problems related to regression analysis in compilation and computation of study material [e.g. comprehensive explorative data analysis (Tukey 1977) before multivariate analyses, data not shown] and by using generalization of the least-square linear regression method (i.e. GLM). Moreover, we used widely accepted model calibration and evaluation procedures (e.g. Beelen et al. 2013) and the HP method for considering potential multicollinearity problems in multivariate analysis (Chevan and Sutherland 1991).

For future studies focusing on spatial prediction of pollen concentrations for exposure assessments, we recommend the following steps (Elith and Leathwick 2009; Beelen et al. 2013): (i) establishment of a conceptual model that is based on literature and empirical findings, (ii) compilation of environmental data from different sources and at various scales (remote sensing-based at finer and land use at coarser scales), (iii) comprehensive statistical and graphical exploration of both response and environmental variable data, and (iv) substantial evaluation of the generated model. The evaluation should include the assessment of the realism of fitted explanatory variables (e.g. expected signs for regression coefficients), the model's fit to data, characteristics of residuals, predictive performance on independent test data, and visual/graphical exploration of the predictions.

Conclusions

In this study, we have elaborated the relationship between environmental determinants and atmospheric allergenic pollen concentrations in an urban area. Previous studies have not explored the possibility to predict the intra-urban variation of grass pollen concentrations, using geospatial data and statistical methods. A novelty of the present work is a comprehensive set of pollen measurements in urban environments, which enable the spatial modelling of urban pollen concentrations. Moreover, we developed the LUR approach by exploring the contributions of different data sources (remote sensing and land use) and scales of explanatory variables.

Based on the results, we draw three main conclusions. Firstly, it is possible to spatially predict the fine-scale variation of grass pollen concentrations across an urban area using the LUR approach. An extensive evaluation of the modeling results is highly important. Moreover, based on the visual exploration of pollen predictions (Fig. 4), models should be extrapolated beyond the calibration environment with care. Secondly, a remote sensing vegetation variable (TC greenness) outperformed land use variables in our study setting. Remote sensing-based indices have several strengths, which highlights their utilization in modeling and predicting pollen concentrations in human-modified environments. Thirdly, statistically-based predictive pollen models could probably be used in retrospective exposure assessment, if residential histories are available and pollen concentrations have been monitored or modeled for the corresponding period. The developed LUR approach demonstrated the possibility of predicting the spatial variability of mean pollen concentrations at breathing-zone level, contrary to urban-background or regional scale usually pursued by the existing monitoring and modeling tools.

In future, exposure assessment studies should not be solely based on (few) roof-level pollen monitoring sites due to the significant intra-urban variation of allergenic pollen concentrations. Instead, it would be highly valuable to combine both local and regional scale observations in studying spatially and temporally the relations between environmental determinants and concentrations of allergenic pollen. Moreover, hybrid modeling should be examined that combines physically- (i.e., deterministic dispersion modeling) and statistically-based models.

References

Aina R, Asero R, Ghiani A, Marconi G, Albertini E, Citterio S. 2010. Exposure to cadmium-contaminated soils increases allergenicity of Poa annua L. pollen. Allergy 65:1313–1321.

Aït-Khaled N, Pearce N, Anderson HR, Ellwood P, Montefort S, Shah J, et al. 2009. Global map of the prevalence of symptoms of rhinoconjunctivitis in children: The International Study of Asthma and Allergies in Childhood (ISAAC) Phase Three. Allergy 64:123–148.

Asher MI, Montefort S, Björkstén B, Lai CKW, Strachan DP, Weiland SK, et al. 2006. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. Lancet 368:733–743.

Beelen RM, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, et al. 2013. Development of NO_2 and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – the ESCAPE project. Atmos Environ 72:10–23.

Beggs PJ. 2004. Impacts of climate change on aeroallergens: past and future. Clin Exp Allergy 34:1507–1513.

Beggs PJ, Bambrick HJ. 2005. Is the global rise of asthma an early impact of anthropogenic climate change? Environ Health Perspect 113:915–919.

Blomme K, Tomassen P, Lapeere H, Huvenne W, Bonny M, Acke F, et al. 2013. Prevalence of allergic sensitization versus allergic rhinitis symptoms in an unselected population. Int Arch Allergy Immunol 160:200–207.

Boutin-Forzano S, Moreau D, Kalaboka S, Gay E, Bonnefoy X, Carrozzi L, et al. 2007. Reported prevalence and co-morbidity of asthma, chronic bronchitis and emphysema: a pan-European estimation. Int J Tuberc Lung Dis 11:695–702.

Burnham KP, Anderson DR. 2004. Multimodel inference understanding AIC and BIC in model selection. Sociol Methods Res 33:261–304.

Caillaud D, Thibaudon M, Martin S, Ségala C, Besancenot JP, Clot B, et al. 2014. Short-term effects of airborne ragweed pollen on clinical symptoms of hay fever in a panel of 30 patients. J Investig Allergol Clin Immunol 24:249–256.

Cecchi L, D'Amato G, Ayres JG, Galan C, Forastiere F, Forsberg B, et al. 2010. Projections of the effects of climate change on allergic asthma: the contribution of aerobiology. Allergy 65:1073–1081.

Chevan A, Sutherland M. 1991. Hierarchical partitioning. Am Stat 45:90–96.

Countries. 2008. ESRI Data & Maps 9.3, courtesy of ArcWorld Supplement.

D'Amato G, Cecchi L, Bonini S, Nunes C, Annesi-Maesano I, Behrendt H, et al. 2007. Allergenic pollen and pollen allergy in Europe. Allergy 62:976–990.

D'Amato G, Cecchi L, D'Amato M, Annesi-Maesano I. 2014. Climate change and respiratory diseases. Eur Respir Rev 23:161–169.

Delfino RJ, Zeiger RS, Seltzer JM, Street DH, McLaren CE. 2002. Association of asthma symptoms with peak particulate air pollution and effect modification by anti-inflammatory medication use. Environ Health Perspect 110:A607–A617.

Elith J, Leathwick JR. 2009. Species distribution models: ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40:677–697.

Finnish Environmental Institute. 2009. CLC2006 Finland. Final technical report, Finnish Environmental Institute, Helsinki, Finland.

Gilmour MI, Jaakkola MS, London SJ, Nel AE, Rogers CA. 2006. How exposure to environmental tobacco smoke, outdoor air pollutants, and increased pollen burdens influences the incidence of asthma. Environ Health Perspect 114:627–633.

Gulliver J, de Hoogh K, Hansell A, Vienneau D. 2013. Development and back-extrapolation of NO_2 land use regression models for historic exposure assessment in Great Britain. Environ Sci Technol 47:7804–7811.

Haberle SG, Bowman DM, Newnham RM, Johnston FH, Beggs PJ, Buters J, et al. 2014. The macroecology of airborne pollen in Australian and New Zealand urban areas. PLoS ONE 9:e97925; doi:10.1371/journal.pone.0097925.

Henneberger PK, Redlich CA, Callahan DB, Harber P, Lemiere C, Martin J, et al. 2011. An official American Thoracic Society statement: work-exacerbated asthma. Am J Respir Crit Care Med 184:368–378.

Hjort J, Suomi J, Käyhkö J. 2011. Spatial prediction of urban–rural temperatures using statistical methods. Theor Appl Climatol 106:139–152.

Hofmann F, Otto M, Wosniok W. 2014. Maize pollen deposition in relation to distance from the nearest pollen source under common cultivation-results of 10 years of monitoring (2001 to 2010). Environ Sci Eur 26: 24.

Jaakkola MS, Nordman H, Piipari R, Uitti J, Laitinen J, Karjalainen A, et al. 2002. Indoor dampness and molds and development of adult-onset asthma: a population-based incident case-control study. Environ Health Perspect 110:543–547.

Katz DS, Carey TS. 2014. Heterogeneity in ragweed pollen exposure is determined by plant composition at small spatial scales. Sci Total Environ 485:435–440.

Kukkonen J, Olsson T, Schultz DM, Baklanov A, Klein T, Miranda AI, et al. 2012. A review of operational, regional-scale, chemical weather forecasting models in Europe. Atmos Chem Phys 12:1–87.

Lillesand TM, Kiefer RW, Chipman JW. 2004. Remote Sensing and Image Interpretation. 5th ed. New York:Wiley.

Lovasi GS, O'Neil-Dunne JP, Lu JW, Sheehan D, Perzanowski MS, MacFaden SW, et al. 2013. Urban tree canopy and asthma, wheeze, rhinitis, and allergic sensitization to tree pollen in a New York City birth cohort. Environ Health Perspect 121:494–500.

Majd A, Chehregani A, Moin M, Gholami M, Kohno S, Nabe T, et al. 2004. The effects of air pollution on structures, proteins and alleregenicity of pollen grains. Aerobiologia 20:111–118.

Majkowska-Wojciechowska B, Pełka J, Korzon L, Kozłowska A, Kaczała M, Jarzebska M, et al. 2007. Prevalence of allergy, patterns of allergic sensitization and allergy risk factors in rural and urban children. Allergy 62:1044–1150.

McCullagh P, Nelder JA. 1989. Generalized Linear Models. 2nd ed. New York:Chapman and Hall.

O'Rourke MK, Lebowitz MD. 1984. A comparison of regional atmospheric pollen with pollen collected at and near homes. Grana 23:55–64.

Pallasaho P, Juusela M, Lindqvist A, Sovijärvi A, Lundbäck B, Rönmark E. 2011. Allergic rhinoconjunctivitis doubles the risk for incident asthma - results from a population study in Helsinki, Finland. Respir Med 105:1449–1456.

Peel RG, Hertel O, Smith M, Kennedy R. 2013. Personal exposure to grass pollen: relating inhaled dose to background concentration. Ann Allergy Asthma Immunol 111:548–554.

Pirinen P, Simola H, Aalto J, Kaukoranta J-P, Karlsson P, Ruuhela R. 2012. Climatological Statistics of Finland 1981–2010. Reports 2012:1. Helsinki, Finland:Finnish Meteorological Institute.

Priftis KN, Anthracopoulos MB, Nikolaou-Papanagiotou A, Matziou V, Paliatsos AG, Tzavelas G, et al. 2007. Increased sensitization in urban vs. rural environment - rural protection or an urban living effect? Pediatr Allergy Immunol 18:209–216.

Rantio-Lehtimäki A, Helander ML, Karhu K. 1992. Does cutting of mugwort stands affect airborne pollen concentrations? Allergy 47:388–390.

Richardson DB, Volkow ND, Kwan MP, Kaplan RM, Goodchild MF, Croyle RT. 2013. Spatial turn in health research. Science 339:1390–1392.

Ring J, Akdis C, Lauener R, Schappi G, Traidl-Hoffmann C, Akdis M, et al. 2014. Global Allergy Forum and Second Davos Declaration 2013 Allergy: Barriers to cure - challenges and actions to be taken. Allergy 69:978–982.

Sikoparija B, Radisic P, Pejak T, Simic S. 2006. Airborne grass and ragweed pollen in the southern Panonnian Valley - consideration of rural and urban environment. Ann Agric Environ Med 13:263–266.

Singer BD, Ziska LH, Frenz DA, Gebhard DE, Straka JG. 2005. Increasing Amb a 1 content in common ragweed (Ambrosia artemisiifolia) pollen as a function of rising atmospheric CO_2 concentration. Funct Plant Biol 32:667–670.

Skjøth C, Ørby PV, Becker T, Geels C, Schlünssen V, Sigsgaard T, et al. 2013. Identifying urban sources as cause of elevated grass pollen concentrations using GIS and remote sensing. Biogeosciences 10:541–554.

SLICES. 2011. SLICES land use. National Land Survey of Finland, Helsinki.

Sofiev M, Bergman KC, eds. 2012. Allergenic Pollen. A Review of the Production, Release, Distribution and Health Impacts. Dordrecht:Springer.

Sofiev M, Siljamo P, Ranta H, Linkosalo T, Jaeger S, Rasmussen A, et al. 2012. A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. Int J Biometeorol 57:45–58.

Sokal RR, Rohlf FJ. 1995. Biometry: The Principles and Practice of Statistics in Biological Research. 3rd ed. New York:Freeman.

Sterling DA, Lewis RD. 1998. Pollen and fungal spores indoor and outdoor of mobile homes. Ann Allergy Asthma Immunol 80:279–285.

Stockwell DR, Peterson AT. 2002. Effects of sample size on accuracy of species distribution models. Ecol Modell 148:1–13.

Taylor PE, Jacobson KW, House JM, Glovsky MM. 2007. Links between pollen, atopy and the asthma epidemic. Int Arch Allergy Immunol 144:162–170.

Tukey JW. 1977. Exploratory Data Analysis. Reading:Addison-Wesley.

White JF, Bernstein DI. 2003. Key pollen allergens in North America. Ann Allergy Asthma Immunol 91:425–435.

WHO (World Health Organization). 2013. Review of Evidence on Health Aspects of Air Pollution – REVIHAAP. First Results. Copenhagen, Denmark:WHO Regional Office for Europe.

Zeldin DC, Eggleston P, Chapman M, Piedimonte G, Renz H, Peden D. 2006. How exposures to biologics influence the induction and incidence of asthma. Environ Health Perspect 114:620–626.

Ziska LH, Gebhard DE, Frenz DA, Faulkner S, Singer BD, Straka JG. 2003. Cities as harbingers of climate change: common ragweed, urbanization, and public health. J Allergy Clin Immunol 111:290–295.

Table 1. Environmental determinants computed to explore the intra-urban variation of grass pollen concentrations. The Landsat and SLICES data were from 2010 (SLICES 2011) and CORINE from 2006 (Finnish Environmental Institute 2009).

Environmental variable	Unit	Description	Source
TC Brightness	Index	Land use/cover with high albedo (overall brightness of the image)	Landsat TM5; http://earthexplorer.usgs.gov
TC Greenness	Index	Amount of photosynthetically-active green vegetation	See above
Wasteland	m^2	Unmanaged grasslands (e.g. meadows and power lines)	SLICES land use classification
Park	m^2	Managed grasslands (e.g. urban parks and sports fields)	SLICES land use classification
Field	m^2	Cultivated fields (e.g. wheat and barley fields)	SLICES land use classification
Deciduous forest	m^2	Broadleaved trees (e.g. birch, alder and maple)	CORINE land cover database
Mixed forest	m^2	Broadleaved trees and conifers (pine and spruce)	CORINE land cover database
Urban land use	m^2	Urban land use classes (e.g. high density residential, commercial, industrial and traffic areas)	SLICES land use classification

TC = Tasseled Cap, TM5 = Thematic Mapper 5.

Table 2. Results of the final generalized linear models (GLMs).

Model calibration	**All samples (n = 16)**	**Extreme values off (n = 14)**
Final GLM ($\alpha+\beta x_i$)	1.287+0.032(Greenness)+7.8^{-6}(Wasteland)	1.179+0.024(Greenness)
AIC	18.2	8.0
D^2	0.785	0.466
Correlation (fitted–obs.)	$r = 0.89$ ($p < 0.001$); $r_s = 0.79$ ($p < 0.001$)	$r = 0.68$ ($p = 0.007$); $r_s = 0.69$ ($p = 0.006$)
Residuals		
Normality (K-S test)	Normally distributed ($p = 0.829$)	Normally distributed ($p = 0.826$)
Homoscedasticity	Acceptable	Acceptable
Spatial autocorrelation	No autocorrelation (Moran's I $p > 0.05$)	No autocorrelation (Moran's I $p > 0.05$)
Cook's distance	One observation > 1	All observations < 0.5

α = intercept; β = regression coefficient; x_i = environmental variable; Greenness = greenness of tasseled cap transformation; Wasteland = unmanaged grassland land use classes; AIC = Akaike's Information Criterion; D^2 = explained deviance (comparable to explained variance in least-square regression); fitted = fitted values; obs. = observed values; r = Pearson's correlation coefficient; r_s = Spearman's rank order correlation coefficient; K-S test = Kolmogorov-Smirnov test.

Figure Legends

Figure 1. Land use of the study area and the location of the monitoring sites for grass pollen (Countries 2008; SLICES 2011). Urban land use refers to high density (e.g. block of flats, commercial and industrial) land use types, parks to managed grasslands, fields to cultivated fields and wasteland to unmanaged grasslands.

Figure 2. An example of the application of the land use regression (LUR) outcomes (e.g. average seasonal concentration of pollen) in long-term and life-time exposure assessments. First, the LUR approach is used to produce a raster model of pollen concentrations in the current environmental conditions ($t_{current}$). Second, the retrospective pollen concentrations ($t_1 \ldots t_n$) are predicted using data describing the past environmental conditions (e.g. historical land use and remote sensing data) (Gulliver et al. 2013). In the estimation of retrospective concentrations, long-term (permanent) pollen collectors are used. Third, geographic information system (GIS) tools are used to compute individual (e.g. cohort members) exposure based on residential history data at applicable spatial resolution (Richardson et al. 2013). Fourth, the results of the previous step are used in allergy and asthma explorations.

Figure 3. Scatter plots of the observed (i.e. measured) and predicted grass pollen concentrations (grains m^{-3}) for the larger (n = 16) (A–B) and smaller (n = 14) (C–D) datasets. In the evaluation setting, the observed concentrations were measured during morning (8:00–11:30, June 27–July 9, 2013) (A and C) and another period (8:00–11:30 and 13:00–16:30, July 10–21, 2013) (B and D). The continuous line show the optimum fit (intercept = 0, slope = 1) and the dashed line the fit to the data. Pearson's correlation coefficients (r) are also shown.

Figure 4. A smoothed prediction map of the grass pollen concentration (grains m^{-3}; afternoon situation during June 27–July 9, 2013) at a 100 m resolution and aerial photographs (A–D) (resolution = 50 cm) representing different environmental conditions within the Helsinki Metropolitan Area, Finland. The predictions were computed using the final generalized linear model and n = 14 dataset. To consider the potential statistical problems related to outlier observations in the n = 16 dataset (see the Method and Results sections) the smaller dataset was used in the prediction (i.e. in the extrapolation of the model beyond the environmental conditions of the calibration area).

Figure 1.

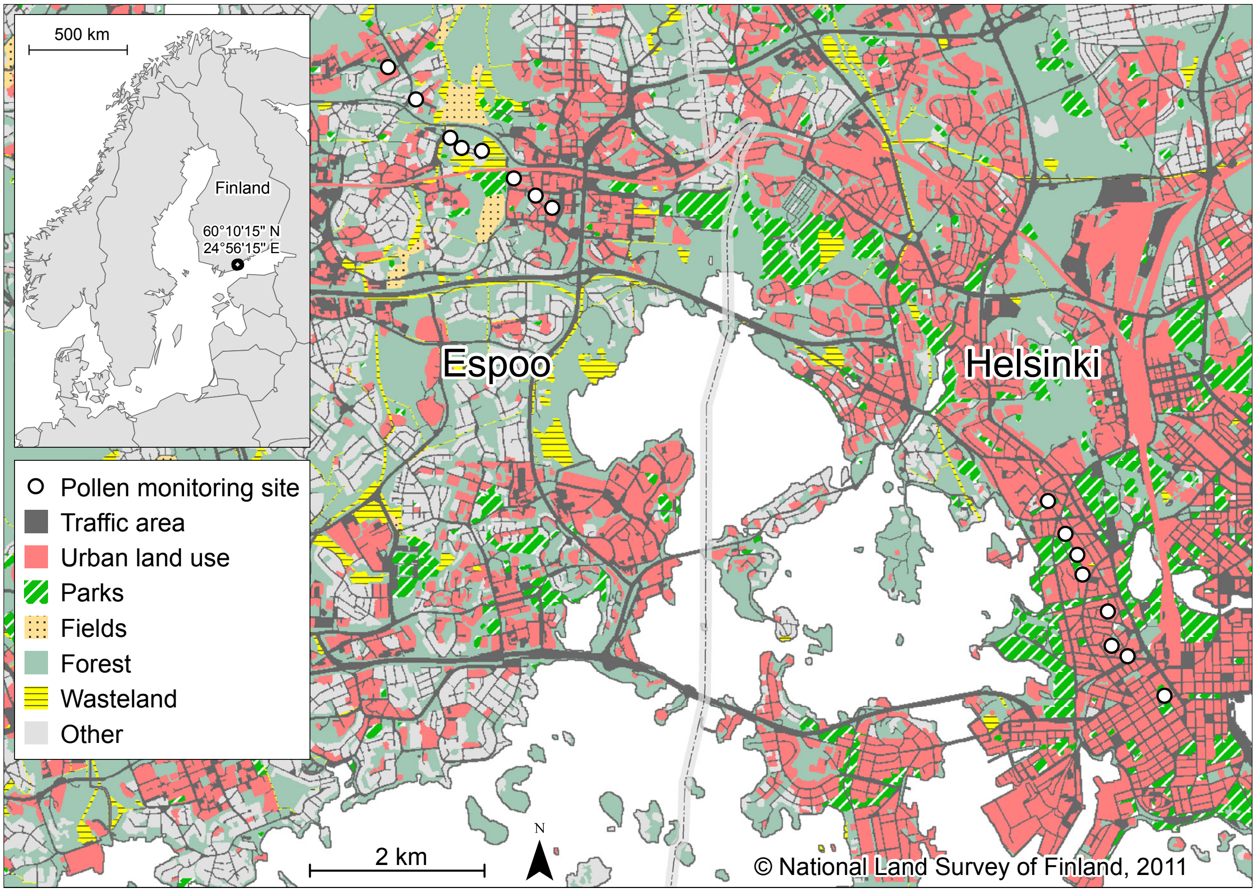

500 km
Finland
60°10'15" N
24°56'15" E
Pollen monitoring site
Traffic area
Urban land use
Parks
Fields
Forest
Wasteland
Other
Espoo
Helsinki
2 km
N
© National Land Survey of Finland, 2011

Figure 2.

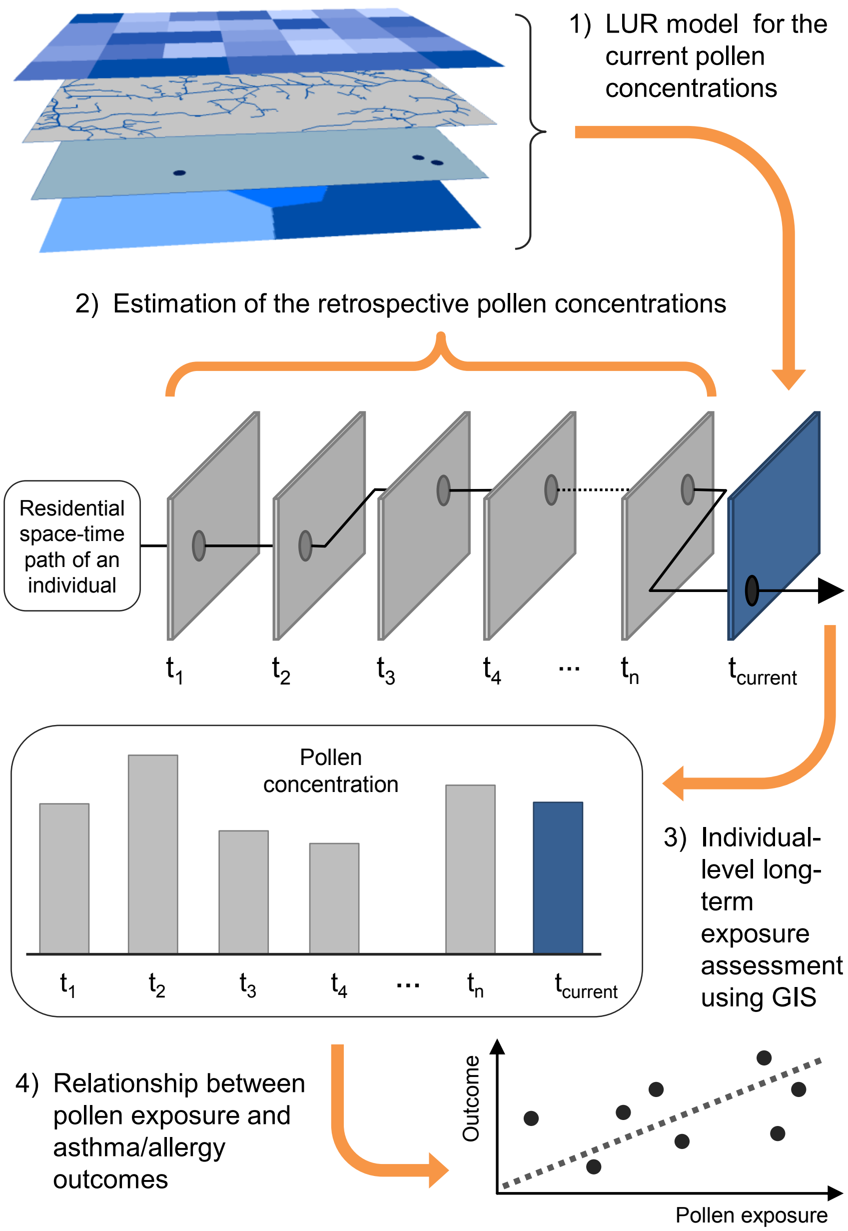

1) LUR model for the current pollen concentrations
2) Estimation of the retrospective pollen concentrations
Residential space-time path of an individual
t1
t2
t3
t4
…
tn
tcurrent
Pollen concentration
t1
t2
t3
t4
…
tn
tcurrent
3) Individual-level long-term exposure assessment using GIS
4) Relationship between pollen exposure and asthma/allergy outcomes
Outcome
Pollen exposure

Figure 3.

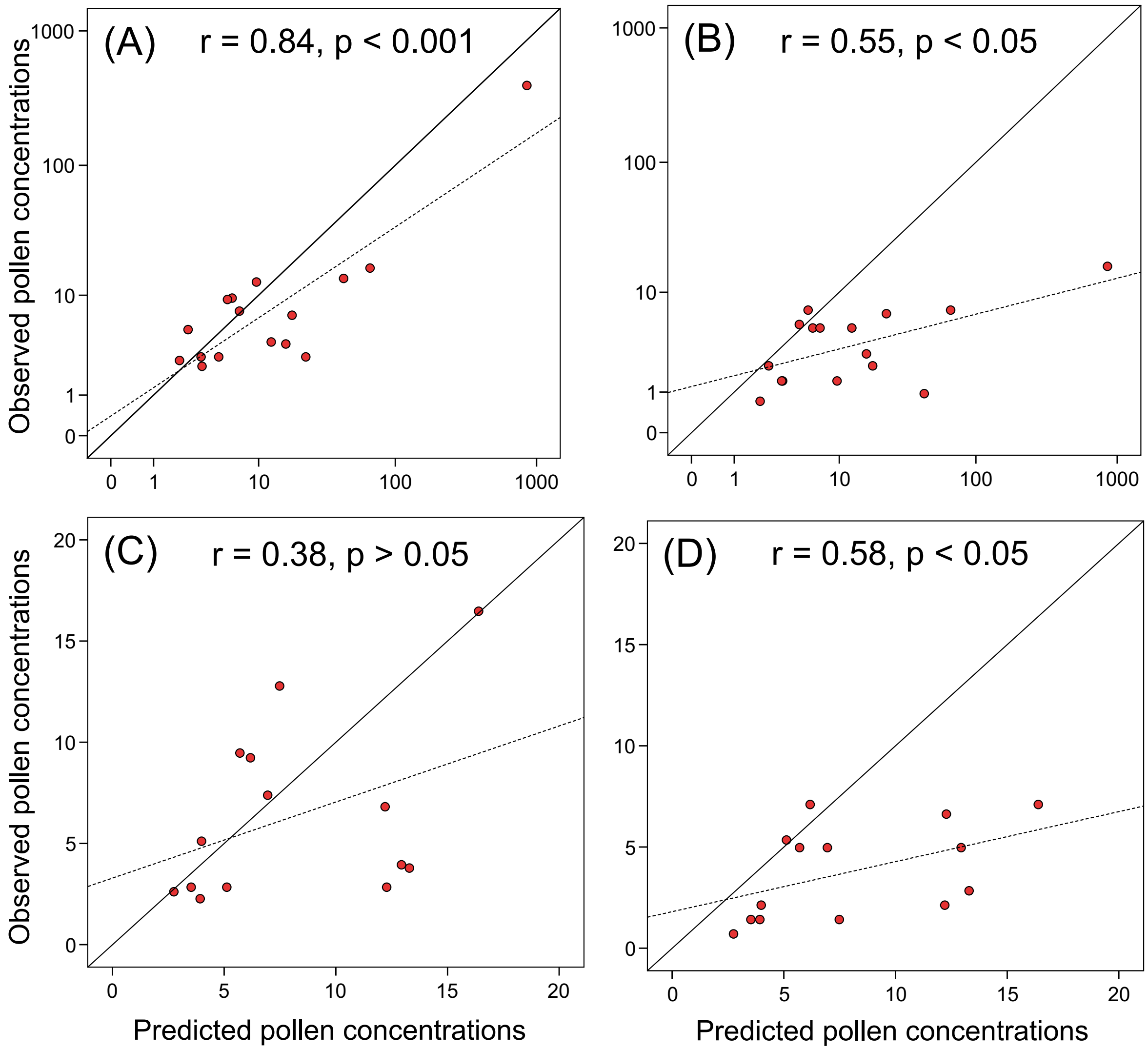
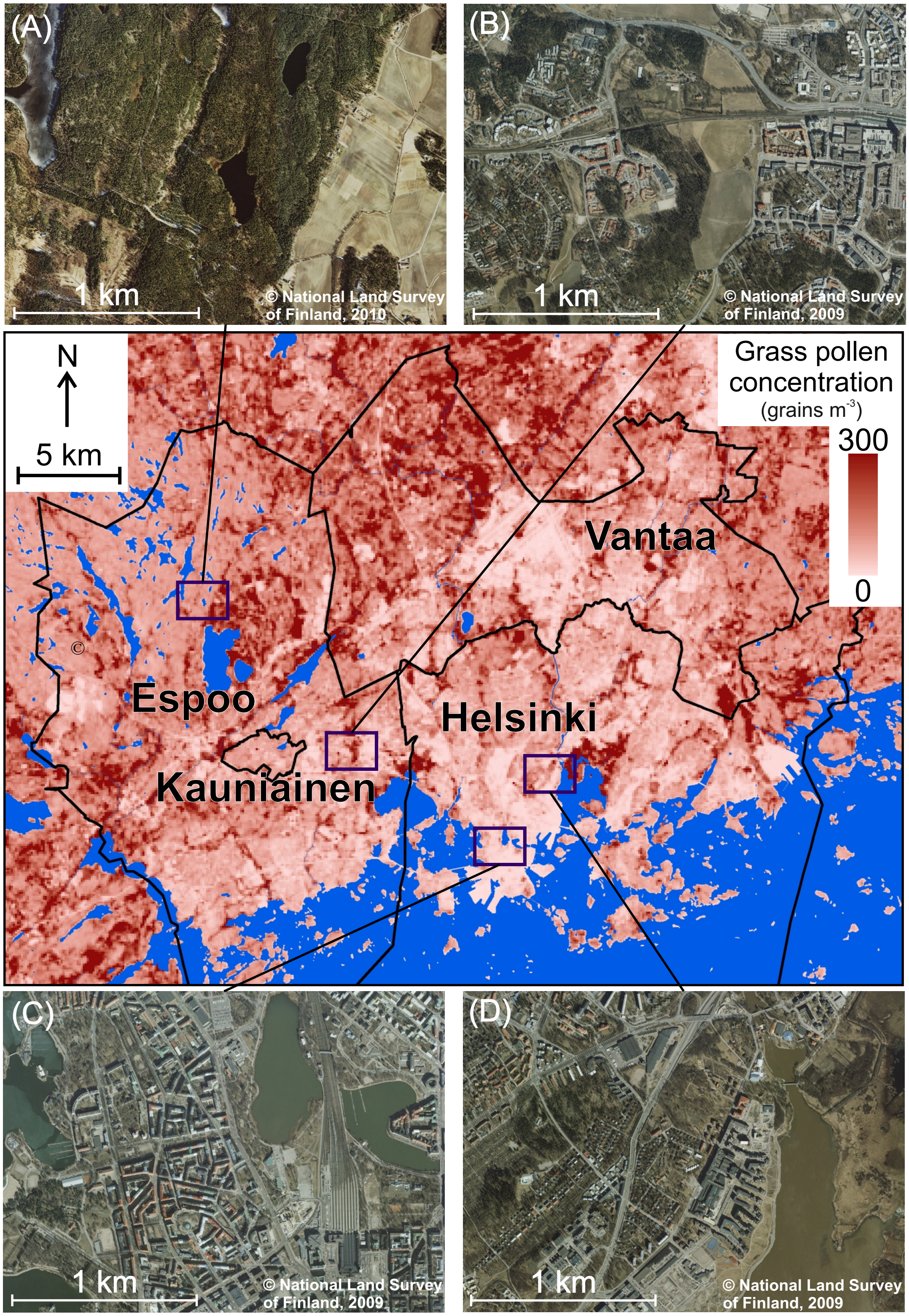

(A) r = 0.84, p < 0.001
(B) r = 0.55, p < 0.05
(C) r = 0.38, p > 0.05
(D) r = 0.58, p < 0.05
Observed pollen concentrations
Observed pollen concentrations
Predicted pollen concentrations
Predicted pollen concentrations
0
1
10
100
1000
0
5
10
15
20

Figure 4.


(A)
1 km
© National Land Survey of Finland, 2010
(B)
1 km
© National Land Survey of Finland, 2009
N
5 km
Grass pollen concentration
(grains m-3)
300
0
Vantaa
Espoo
Helsinki
Kauniainen
(C)
1 km
© National Land Survey of Finland, 2009
(D)
1 km
© National Land Survey of Finland, 2009